



IA agéntica en educación superior

Fundamentos, arquitecturas, diseño pedagógico y gobernanza de los agentes en la formación universitaria

Fernando Santamaría

PRIMERA EDICIÓN · 2026

IA agéntica en educación superior

Fundamentos, arquitecturas, diseño pedagógico y gobernanza de los agentes en la formación universitaria.

*La autonomía operativa se delega.
La agencia epistémica, no.*

QUÉ ENCONTRARÁ AQUÍ

Los preliminares completos, el capítulo 1 íntegro con sus cinco figuras y el apartado 2.6, el que sostiene la tesis. Unas 7.000 palabras de las cerca de 85.000 de la obra, más cuatro tablas de síntesis compuestas para esta muestra.

QUÉ VIVE EN EL LIBRO

Los instrumentos de trabajo: el marco de autonomía pedagógicamente restringida con sus cinco niveles, la matriz de decisión, la plantilla de evaluación de herramientas y el canvas de diseño. En total, 20 capítulos, 7 anexos y 73 figuras.

ARQUITECTURA COMPLETA DE LA OBRA · SEIS PARTES

I Fundamentos conceptuales y ontológicos

Qué separa a un agente de un sistema conversacional y qué teorías del aprendizaje sostienen su uso.

II Arquitecturas, patrones y coordinación

Anatomía de un sistema agéntico educativo y patrones de diseño cuando trabajan varios agentes a la vez.

III Diseño pedagógico con agentes

Roles pedagógicos, marco de autonomía pedagógicamente restringida y diseño de instrucciones.

IV Aplicaciones en educación superior

Tutoría, casos documentados y verificados, y agentes al servicio de la investigación académica.

V Confiabilidad, evaluación y gobernanza

Cómo saber si un agente funciona, seguridad sistémica, ética y gobernanza institucional.

VI Perspectivas y recomendaciones

Hacia una pedagogía agéntica reflexiva y líneas de investigación emergentes.

IA agéntica en educación superior. Fundamentos, arquitecturas, diseño pedagógico y gobernanza de los agentes en la formación universitaria.

Fernando Santamaría González · Editorial IAforTeachers · Primera edición, 2026

Muestra gratuita, de distribución libre. La obra completa se publica el jueves 3 de septiembre de 2026.

Índice

ANTES DE EMPEZAR

Preliminares	05
Introducción	05
Nota sobre el método	07
Sobre el autor	08

PARTE I · FUNDAMENTOS CONCEPTUALES Y ONTOLÓGICOS

1	De la IA conversacional a la IA agéntica	10
1.1	Qué aporta la IA agéntica al aprendizaje universitario: más allá de la asistencia conversacional	10
1.2	Diferencias con la IA educativa tradicional: del asistente conversacional al sistema autónomo	11
1.3	Las cuatro capacidades distintivas de los agentes educativos	13
1.4	El bucle agéntico: percibir, planificar, actuar y evaluar	14
1.5	De los sistemas basados en reglas a las arquitecturas agénticas: qué transición estamos viviendo	16
1.6	Clarificación terminológica: agente, agenticidad, IA agéntica	17
2	El problema ontológico del agente	19
2.6	Autonomía operativa y agencia epistémica: qué puede delegarse y qué no	19

APARATO

Referencias	23
-------------------	----

PARTE I

Fundamentos conceptuales y ontológicos

Qué separa a un agente de un sistema conversacional, de dónde viene la palabra agente y qué teorías del aprendizaje sostienen su uso.

-
- 1 De la IA conversacional a la IA agéntica
 - 2 El problema ontológico del agente
 - 3 Bases teóricas del aprendizaje mediado por agentes

ANTES DE EMPEZAR

Preliminares



Prólogo, introducción, nota sobre el método y sobre el autor.

Los agentes de inteligencia artificial han entrado en las aulas universitarias antes de que dispongamos de un vocabulario común para nombrarlos, de evidencia suficiente para juzgarlos y de acuerdos para gobernarlos. La promesa suena alta: un tutor incansable, retroalimentación inmediata, personalización a escala. Este libro baja el volumen de ese estruendo. **Hemos escrito una guía para pensar con criterio mientras la tecnología avanza y la evidencia trata de alcanzarla.**

El avance técnico de los agentes ha corrido más que la evaluación, la seguridad y la gobernanza. Algunas aplicaciones ya resultan útiles. Como ingeniería confiable, en cambio, el campo sigue siendo joven. Esa tensión vertebra el volumen y explica una decisión metodológica que hemos mantenido con obstinación: distinguir lo documentado de lo emergente, y lo plausible de lo especulativo. Cuando hay evidencia, la citamos. Cuando falta, lo decimos.

Este manual no enseña a programar ni celebra la herramienta de la semana. Tampoco condena una tecnología que, bien usada, puede enriquecer el aprendizaje. Las nociones de arquitectura, coordinación y seguridad aparecen en la medida en que ayudan a decidir. El destinatario es el profesorado universitario hispanohablante que ya convive con estos sistemas, a veces sin haberlos elegido, y necesita conceptos, criterios y cautelas para actuar.

Todos los capítulos vuelven a una misma frontera: la que separa la autonomía operativa de la agencia epistémica. Podemos delegar en un sistema la ejecución de ciertas tareas, incluso conviene hacerlo en algunos casos. La capacidad de construir juicio propio, comprender razones y responder por una conclusión pertenece a la persona. **Si cedemos esa capacidad, vaciamos de sentido el acto de educar.** La evaluación, la retroalimentación, el papel docente, la equidad y la responsabilidad son variaciones de esta pregunta: qué cedemos a la máquina y qué decidimos preservar.

Escribimos este libro desde una convicción y una cautela. La convicción de que los agentes pueden amplificar el aprendizaje si se diseñan con inteligencia pedagógica. La cautela de que, mal empleados, pueden atrofiar justamente aquello que dicen potenciar. Entre esas dos fuerzas se juega buena parte del futuro de la docencia universitaria, y ninguna herramienta lo decidirá por nosotros. Lo decidiremos con las opciones que tomemos cada día en el aula. A ayudar a tomarlas mejor se dedican las páginas que siguen.

Introducción

Un chatbot espera una pregunta y responde. Un agente puede percibir un entorno, planificar varios pasos, recurrir a herramientas externas y revisar lo que ha hecho antes de continuar. Esa capacidad de actuar con cierta autonomía introduce posibilidades que la IA conversacional no ofrecía y riesgos que tampoco planteaba. Ahí empieza este manual.

Todo libro sobre esta materia afronta la tentación de confundir lo posible con lo probado. La literatura sobre IA educativa abunda en promesas que la evidencia todavía no respalda. Para mantener las categorías separadas, aplicamos una escala de madurez a cada afirmación relevante. Lo documentado cuenta con evidencia pública verificable. Lo emergente existe, aunque su alcance sigue sin estabilizarse. Lo plausible es técnicamente posible y aún no se ha demostrado. Lo especulativo proyecta un desarrollo razonable que puede no cumplirse. **La escala indica al lector sobre qué terreno pisa.**

De ese rigor se desprende una decisión que algunos podrían juzgar excesiva. Solo tres casos merecen, a nuestro juicio, la etiqueta de plenamente documentados: Khanmigo, de Khan Academy, Pounce, de Georgia State University, y PitchQuest, de la Wharton School. Los tres aparecen analizados con detalle, y con una observación que atraviesa el libro. Existe una relación inversa entre la ambición pedagógica de una intervención y la solidez de su evidencia. La aplicación más acotada, Pounce, es la que dispone de ensayos controlados. La más ambiciosa descansa sobre un artículo de diseño. Conviene recordarlo cada vez que alguien anuncia una revolución.

La distinción entre autonomía operativa y agencia epistémica resume la posición del manual. La primera designa lo que un agente puede hacer por su cuenta: ejecutar, calcular, redactar un borrador u organizar información. Delegar algunas de esas tareas libera tiempo y energía. La segunda pertenece al estudiante: construir conocimiento, sopesar razones, dudar y decidir. Su delegación desactiva el aprendizaje. Los debates sobre evaluación, retroalimentación, personalización y gobierno se ordenan alrededor de esta frontera. La pregunta pedagógica es cuánto debe reservarse para la persona, aunque la máquina pueda hacer más.

Para pensar esa frontera con precisión, hemos desarrollado un conjunto de marcos propios que el lector irá encontrando. El más estructural es el marco de *autonomía pedagógicamente restringida*, que gradúa en cinco niveles, de APR-0 a APR-4, cuánta autonomía conviene conceder a un agente según la tarea y el estudiante. A su alrededor orbitan nociones que dan nombre a lo que antes no lo tenía. La zona de desarrollo coagéntico reformula a Vygotsky para pensar qué alcanza el estudiante con el apoyo del agente. El provocador epistémico describe a quien devuelve la pregunta en lugar de la respuesta. **Y la presencialidad cognitiva nombra la implicación de un pensamiento humano que acompaña y sostiene el de otro.**

La *coagencia* también tiene trampas. Llamamos *ilusión agéntica* a la sensación de haber aprendido cuando apenas se ha completado una tarea. La descarga cognitiva puede liberar la mente o atrofiar una capacidad que deja de ejercitarse. Y la *brecha agéntica* separa a quien usa el sistema para ampliar su juicio de quien acaba sustituyéndolo. Proponemos dos respuestas concretas: la *desconexión agéntica programada*, que reserva espacios de esfuerzo personal, y el marco de *transparencia coagéntica*, que permite saber quién hizo qué. El diseño decide si la colaboración amplía una capacidad o la empobrece.

El recorrido se organiza en seis partes. La primera asienta los fundamentos conceptuales y ontológicos, qué es un agente y qué teorías del aprendizaje nos ayudan a pensarlo. La segunda abre la caja de las arquitecturas, los patrones de diseño y la coordinación entre varios agentes, siempre al nivel que un docente necesita. La tercera, núcleo pedagógico de la obra, presenta los roles de los agentes, el marco APR y el diseño de instrucciones. La cuarta baja al terreno de las aplicaciones, con los casos documentados, las posibilidades plausibles y la línea emergente de la investigación asistida. La quinta aborda la confiabilidad, la evaluación, la seguridad y la gobernanza, el frente donde el campo todavía es joven. La sexta ofrece perspectivas y recomendaciones para instituciones y docentes.

Este manual está pensado para quien debe decidir. Puede leerse de principio a fin, siguiendo el arco que va del concepto a la práctica, o consultarse por partes. Una pregunta acompaña ambos recorridos: ¿el uso que hacemos de estos agentes construye nuestra capacidad de pensar y la de nuestros estudiantes, o la sustituye? Cada institución y cada docente tendrá que responderla. **El libro ofrece criterios para hacerlo.**

Nota sobre el método

Un libro sobre agentes de inteligencia artificial tiene una obligación que otros no tienen, la de decir cómo se escribió. Escribo esta nota porque la afirmación de que un texto se ha redactado a mano vale poco si no se especifica dónde no se hizo así, y porque un lector que va a fiarse de doscientas sesenta y cinco páginas sobre esta materia merece saber en qué punto intervino la máquina.

La versión original del texto es mía. También lo son la tesis que lo vertebra, la escala de madurez, el marco APR, los conceptos propios y los juicios sobre la calidad de la evidencia. Respondo de esas decisiones y de los errores que queden.

En agosto de 2026 sometí la obra completa a una revisión editorial asistida por inteligencia artificial. La intervención se concentró en el estilo: eliminar repeticiones, construcciones formularias, calcos del inglés y ritmos demasiado regulares. La revisión no incorporó casos, datos, referencias ni conclusiones nuevas. Cuando una frase afectaba al argumento o al grado de certeza de una afirmación, la decisión final fue autoral.

Este libro se ha construido despacio, paso a paso, y las herramientas se emplearon en tres momentos de ese recorrido, todos ellos de búsqueda y contraste.

Para localizar literatura utilicé Elicit, Consensus y Perplexity, que permiten llegar a trabajos que los buscadores académicos convencionales no devuelven en un campo que publica a esta velocidad. Ninguna referencia entró en este libro por haber sido sugerida. **Todas se leyeron antes de citarse, y las que no resistieron la lectura se quedaron fuera.**

Para analizar los documentos ya localizados recurrí a SciSpace, Scite y Semantic Scholar. El segundo merece una mención particular, porque muestra si un trabajo posterior apoya, discute o solo menciona al que cita, y esa distinción es la que separa una cita real de una cita de adorno. Para cartografiar el campo y encontrar lo que faltaba usé Research Rabbit y Connected Papers, que representan las relaciones entre trabajos y hacen visibles los huecos de una bibliografía.

Para trabajar sobre mis propias fuentes empleé NotebookLM, que solo responde sobre los documentos que uno le entrega y por tanto no puede inventarse lo que no está.

Para verificar el aparato bibliográfico completo utilicé Claude, con consultas directas a los registros oficiales de Crossref y DataCite. Las ciento noventa y seis entradas de este manual se cotejaron una a una, comprobando que el identificador resuelve y que el título, los autores, el año, el volumen y las páginas coinciden con lo declarado. El resultado puede comprobarse referencia por referencia.

Esa verificación produjo ocho correcciones y conviene declararlo. Siete fueron de precisión, y la más significativa afectaba a una cifra clásica que el libro daba sin el ajuste que la investigación posterior le hizo. Corregirlas exigió reescribir algunas frases, de modo que hay en este manual un puñado de oraciones cuya redacción final salió de ese proceso de revisión. Son ocho pasajes localizados y están documentados, no un texto generado.

La revisión no sustituyó la lectura de las fuentes ni delegó el juicio sobre su calidad. Ningún artículo se citó a partir de un resumen automático. **La clasificación de cada caso como documentado, emergente, plausible o especulativo siguió siendo una decisión mía, tomada capítulo por capítulo.**

Sostengo que esta manera de trabajar es la que este libro predica. Sería incoherente defender que la capacidad de comprobar no debe delegarse nunca y haber delegado precisamente eso al escribirlo.

Sobre el autor

Soy filólogo de formación, licenciado en Filología Hispánica por la Universidad de León, y completé después el Diploma de Estudios Avanzados en Psicología y Ciencias de la Educación, con una investigación sobre la implementación de las aplicaciones web de nueva generación en la formación de maestros. De ahí en adelante, veinte años en tecnología educativa.

El recorrido tiene una lógica que conviene contar, porque explica desde dónde está escrito este libro. Empecé con la web social entre 2006 y 2010, cuando casi ninguna universidad española la miraba. Escribí entonces el prólogo a la edición en español de *Conociendo el conocimiento*, de George Siemens, que fue por donde el conectivismo entró en la conversación académica hispanohablante. Seguí con las comunidades de aprendizaje y los entornos personales, después con analítica del aprendizaje, y hoy con inteligencia artificial. **El hilo es siempre el mismo: qué le hace una tecnología a la capacidad de pensar de quien aprende.** Este manual es la última forma de esa pregunta.

Buena parte de mi trabajo ha ocurrido al otro lado del Atlántico. He dado clase en varias universidades colombianas, entre ellas Uniminuto en Bogotá y la Universidad de La Sabana, soy profesor visitante de la Universidad del Salvador (USAL) de Buenos Aires desde 2014 y trabajé como asesor en asuntos tecnológicos en Bogotá. He dado conferencias en Colombia, México, Ecuador, Nicaragua, Argentina, Panamá y España. Desde 2010 tutorizo la cátedra de ecosistemas digitales en la maestría de entornos virtuales de Aprende Virtual, que es donde he acumulado la mayor parte de mi experiencia formando profesorado universitario. **Los ejemplos de este libro salen de esas aulas, no de un despacho.**

En estos veinte años he visto pasar unas cuantas corrientes de entusiasmo tecnológico. Cada una llegó con su promesa de transformar la enseñanza y casi todas dejaron las aulas más o menos como estaban. La razón rara vez fue técnica. **Fue que nadie se molestó en distinguir lo que estaba probado de lo que solo sonaba bien.**

Con los agentes puede que no ocurra lo mismo, y conviene explicar por qué. La diferencia no es de grado. Un sistema que se limita a responder deja intacto el reparto del trabajo docente, porque alguien tiene que preguntar primero. **Un sistema que planifica, recuerda y actúa por su cuenta toca ese reparto.** Ahí hay un cambio real, y por eso mismo hace falta más cautela, no menos.

Escribí este manual porque veía un hueco. Se habla mucho de agentes y falta una estructura con la que pensarlos. Quise construir ese marco y darle prioridad a lo práctico, de modo que el profesorado, y también quien no da clase pero trabaja en esto, pueda entrar en un campo con mucho futuro por delante **sin tener que crearme a mí.** De ahí la escala de madurez que atraviesa estas páginas, y de ahí que verificara las referencias una a una.

Dirijo la Editorial IAforTeachers, donde publiqué hace un mes «La evaluación del aprendizaje en tiempos de la IA. Manual de referencia para rediseñar, integrar y gobernar la evaluación en la educación superior». Es el espacio desde el que seguiré publicando, y habrá más títulos sobre inteligencia artificial con agentes.

Este libro se cierra en 2026 y el campo no. Lo que venga después lo iré publicando en blog.iaforteachers.com. Ahí seguirán los artículos sobre agentes, con el mismo criterio que atraviesa estas páginas: lo que la evidencia acabe confirmando, lo que la desmienta y lo que siga sin estar probado.

Un apunte sobre el ritmo. Este manual son meses de trabajo, y solo en el último mes he tenido que rehacer cosas que ya daba por cerradas. **El campo avanza a una velocidad que ningún lector normal, y me incluyo, puede abarcar entera.** Por eso agradezco las correcciones más que los elogios.

FERNANDO SANTAMARÍA GONZÁLEZ · EN BREVE

FORMACIÓN

Filología Hispánica, Universidad de León. Diploma de Estudios Avanzados en Psicología y Ciencias de la Educación.

DOCENCIA

Profesor visitante de la Universidad del Salvador (USAL), Buenos Aires, desde 2014. Cátedra de ecosistemas digitales en la maestría de entornos virtuales de Aprende Virtual desde 2010. Uniminuto y Universidad de La Sabana, en Colombia.

CONTRIBUCIÓN DE REFERENCIA

Prólogo a la edición en español de *Conociendo el conocimiento*, de George Siemens, texto con el que el conectivismo entró en la conversación académica hispanohablante.

ESCRITURA CONTINUADA

fernandosantamaria.com desde 2003, veintitrés años de escritura pública sobre ecologías del aprendizaje. Blog editorial en blog.iaforteachers.com.

Del mismo autor, ya disponible: «La evaluación del aprendizaje en tiempos de la IA. Manual de referencia para rediseñar, integrar y gobernar la evaluación en la educación superior». Editorial IAforTeachers.

1

De la IA conversacional a la IA agéntica

CAPÍTULO INCLUIDO ÍNTEGRO

Durante casi tres años, la conversación sobre inteligencia artificial en la universidad giró en torno a una misma pregunta. ¿Qué puede responder el modelo? Profesorado y estudiantado aprendieron a formular instrucciones, a depurar respuestas y a convivir con un interlocutor que, por elocuente que pareciera, seguía dependiendo de que alguien iniciara cada intercambio. **Ese régimen conversacional, cómodo y previsible, empieza a resultar insuficiente.**

Un tipo de sistema distinto ha entrado en las aulas sin hacer demasiado ruido. No espera una pregunta para actuar. Planifica secuencias de varios pasos, recurre a herramientas externas y conserva memoria de interacciones anteriores. Lo llamamos IA agéntica. Antes de discutir sus usos educativos necesitamos precisar qué hace, en qué se diferencia de la IA conversacional y qué ciclo operativo sostiene su conducta. La terminología importa, porque buena parte de la confusión comercial empieza en los nombres.

Una advertencia sobre el aparato de este capítulo. Buena parte de lo que hoy se publica sobre agentes educativos son propuestas recientes con escasa acumulación de citas. Hemos preferido apoyar cada afirmación en el trabajo más consolidado disponible para cada asunto, distinguiendo la evidencia técnica sobre agentes, que ya cuenta con revisiones sistemáticas amplias, de la evidencia pedagógica sobre su uso en educación superior, mucho más escasa. Cuando solo disponemos de lo segundo en forma preliminar, lo decimos.

1.1 Qué aporta la IA agéntica al aprendizaje universitario: más allá de la asistencia conversacional

La universidad no llega virgen a esta discusión. Laato y colaboradores (2023) fueron de los primeros en formular, para un público universitario, qué implicaba que el estudiantado dispusiera de un generador de texto competente y disponible, y advirtieron ya entonces que la respuesta institucional no podía reducirse a la prohibición ni a la adopción acrítica. Eager y Brunton (2023) desplazaron el foco desde la herramienta hacia la práctica docente, proponiendo un modelo de enseñanza aumentada donde el sistema amplía el repertorio del profesorado sin ocupar su lugar. La revisión de Yigci y colaboradores (2025) sobre asistentes conversacionales basados en modelos de lenguaje en educación superior recoge el balance de ese primer periodo, con beneficios documentados en accesibilidad y disponibilidad, y limitaciones persistentes en exactitud, sesgo y verificación.

Lo que llega ahora es de otro orden. Un asistente conversacional responde y se retira. Pensemos en un ejemplo concreto. Un estudiante de segundo curso pide que le expliquen la diferencia entre validez interna y validez externa en un diseño experimental. **El asistente devuelve una explicación razonable y la interacción termina ahí.** Un agente educativo podría, en cambio, sostener un objetivo a lo largo del semestre, advertir que ese mismo concepto reaparece mal aplicado en tres entregas sucesivas y proponer, sin que nadie se lo pida, una secuencia de ejercicios dirigida a esa laguna concreta.

¿Aporta esto algo que la buena docencia presencial no ofreciera ya? Probablemente no en esencia, pero sí en escala. Un docente con doscientos estudiantes matriculados puede, en principio, ofrecer un seguimiento longitudinal que antes solo era viable en grupos reducidos. Ahora bien, la escala tiene un precio cognitivo que conviene anticipar. Essel y colaboradores (2023) observaron, en estudiantado de grado que recibía respuestas inmediatas de un modelo de lenguaje, efectos desiguales sobre las competencias de orden superior, con mejoras en algunas tareas y sin mejora apreciable en otras. La disponibilidad permanente de una respuesta correcta no produce por sí sola mejor razonamiento.

Sobre la categoría misma de agente educativo la literatura empieza a ordenarse. Kostopoulos y colaboradores (2025) sitúan la IA agéntica como una clase cualitativamente distinta de la inteligencia artificial educativa previa, capaz de perseguir objetivos complejos y de larga duración con intervención humana mínima. Córdova-Esparza (2025), tras revisar sistemáticamente ochenta y dos estudios sobre agentes educativos publicados entre enero de 2023 y febrero de 2025, encuentra un patrón que conviene tomar en serio, los flujos híbridos en los que el profesorado cura y modera la salida del modelo superan a los tutores completamente autónomos. Ambos trabajos son recientes y aún poco citados, de modo que los tratamos como orientación razonada y no como evidencia asentada.

De ahí se sigue la salvedad epistémica que atraviesa el manual entero. Lo que la investigación técnica sobre agentes documenta con solidez creciente no coincide con lo que la investigación educativa ha verificado en condiciones de aula real. Sostenemos, y lo justificamos en el capítulo 11 con un análisis pormenorizado de tres casos, que la evidencia empírica en educación superior sigue siendo escasa frente a la evidencia técnica general. Khanmigo, de Khan Academy, ilustra ya hoy una aplicación con evidencia documentada. Constituye la excepción.

1.2 Diferencias con la IA educativa tradicional: del asistente conversacional al sistema autónomo

Conviene precisar qué entendemos por IA educativa tradicional, porque el término asistente conversacional se ha usado durante años para designar realidades muy distintas entre sí. Los sistemas de tutoría inteligente clásicos, como AutoTutor (Graesser et al., 2005), ya incorporaban adaptación al ritmo del estudiante y diálogo en lenguaje natural. Operaban, con todo, dentro de reglas fijas, con rutas de decisión predefinidas por quien los programó, y su personalización se limitaba a ajustar la dificultad o la secuencia dentro de un guion cerrado. La revisión de Yigci y colaboradores (2025) muestra que la primera generación de asistentes basados en modelos de lenguaje heredó esa limitación por otra vía, la de carecer de memoria entre sesiones y de iniciativa propia.

¿Dónde trazar entonces la frontera? Kostopoulos y colaboradores (2025) proponen un criterio operativo. Un sistema se considera agéntico cuando satisface al menos cuatro de seis condiciones verificables, entre ellas la planificación de varios turnos, la memoria persistente del progreso, la iniciativa autónoma para lanzar una intervención y la orientación hacia un objetivo sostenido en el tiempo. **La propuesta resulta útil por su carácter comprobable, aunque no está consolidada en la literatura.** Las taxonomías técnicas más recientes coinciden en el fondo y difieren en la forma, y describen la agenticidad como una gradación de arquitecturas antes que como un umbral (Xu, 2026, Arunkumar et al., 2026).

La diferencia excede lo técnico, y aquí es donde el matiz pedagógico decide. Un asistente conversacional espera. El estudiante inicia, el sistema responde, la interacción se cierra. Un agente educativo puede iniciar. Detecta un patrón de error recurrente y actúa antes de que el estudiante formule la pregunta. Esta inversión de la iniciativa, de reactiva a proactiva, es probablemente el cambio de mayor calado que la IA agéntica introduce en el aprendizaje universitario. Es también el que exige mayor cautela. Una intervención no solicitada plantea preguntas legítimas sobre la autonomía del estudiante, y las retomaremos en el capítulo 9.

Para quien decide una compra institucional, esta frontera tiene consecuencias inmediatas. Un asistente sin memoria ni iniciativa se evalúa con un banco de preguntas y respuestas. Un agente que planifica, recuerda y actúa exige observar trayectorias completas, no respuestas sueltas, y obliga a preguntar dónde se almacena el historial del estudiante y quién puede consultarlo. El capítulo 15 desarrolla los procedimientos de evaluación que esta diferencia impone.

FIGURA
1.1

CAPÍTULO 1

De la IA conversacional a la IA agéntica

Cinco quiebres de comportamiento cambian la relación entre la persona y el sistema

IA CONVERSACIONAL · RESPONDE

01 · INICIATIVA

De responder a anticiparse

Detecta una necesidad y propone el siguiente paso sin esperar una petición explícita.

02 · CONTINUIDAD

De la sesión al proceso

Conserva contexto útil y reconoce la trayectoria de trabajo a lo largo del tiempo.

03 · PLANIFICACIÓN

De una respuesta a una secuencia

Descompone un objetivo amplio en tareas ordenadas antes de intervenir.

IA AGÉNTICA · ACTÚA CON PROPÓSITO

04 · ACCIÓN

Del texto a las herramientas

Consulta fuentes, utiliza sistemas externos y ejecuta operaciones autorizadas.

05 · CONTROL

Del resultado a la revisión

Evalúa lo obtenido y corrige el curso cuando no satisface el objetivo.

IDEA CLAVE La diferencia no está en conversar mejor. Está en sostener un objetivo y actuar sobre él.

Figura 1.1. De la IA conversacional a la IA agéntica: cinco quiebres de comportamiento. Elaboración propia a partir de Wang et al. (2024) y Kostopoulos et al. (2025).

TABLA 1.1 · ASISTENTE CONVERSACIONAL Y AGENTE EDUCATIVO

Dimensión	Asistente conversacional	Agente educativo
Iniciativa	Espera. El estudiante inicia, el sistema responde y la interacción se cierra.	Puede iniciar. Detecta un patrón de error recurrente y actúa antes de que llegue la pregunta.
Memoria	Sin memoria entre sesiones. Cada conversación empieza <i>de cero</i> .	Memoria persistente del progreso del estudiante.
Planificación	Responde turno a turno, dentro de rutas de decisión predefinidas.	Planifica varios turnos y sostiene un objetivo en el tiempo.
Personalización	Ajusta dificultad o secuencia dentro de un <i>guion cerrado</i> .	Orientada a un objetivo sostenido, no a un guion.
Cómo se evalúa antes de comprarlo	Con un banco de preguntas y respuestas.	Observando trayectorias completas , y preguntando dónde se almacena el historial y quién puede consultarlo.

Elaboración propia a partir de Kostopoulos et al. (2025), Yigci et al. (2025) y Graesser et al. (2005). El criterio de Kostopoulos exige al menos cuatro de seis condiciones verificables, de las que aquí se recogen las cuatro citadas en el texto.

1.3 Las cuatro capacidades distintivas de los agentes educativos

Con variaciones de nomenclatura, la literatura técnica converge en un núcleo de cuatro capacidades que distinguen a un agente de un generador de texto. Wang y colaboradores (2024), en la revisión más citada hasta la fecha sobre agentes autónomos basados en modelos de lenguaje, las sitúan como el núcleo constructivo de cualquier agente. Las taxonomías posteriores reordenan las piezas sin alterar el fondo (Xu, 2026, Arunkumar et al., 2026). La figura 1.2 sintetiza esas cuatro capacidades aplicadas al contexto educativo.

La primera es la autonomía operativa y la toma de decisiones. El agente evalúa opciones y selecciona un curso de acción sin que un humano intervenga en cada paso, en lugar de limitarse a generar la continuación más probable de una entrada (Wang et al., 2024). En evaluación asistida, esto significa que el sistema puede decidir por sí mismo qué rúbrica aplicar a una entrega antes de que el docente revise el resultado.

Viene después la planificación y el razonamiento de varios pasos, la capacidad que más atención ha recibido. Huang y colaboradores (2024) ordenan la investigación disponible en cinco familias, descomposición de tareas, selección entre planes alternativos, recurso a módulos externos, reflexión sobre los propios errores y apoyo en la memoria, y muestran que casi todos los sistemas competentes combinan varias de ellas. Plaat y colaboradores (2026), en una revisión sistemática del razonamiento de varios pasos, documentan tanto las ganancias de la descomposición explícita como su punto débil, la propagación del error entre pasos sucesivos. Masterman y colaboradores (2024) llegan a una conclusión convergente desde el terreno de las arquitecturas. Un agente de apoyo al diseño instruccional que identifica primero los resultados de aprendizaje, busca después actividades coherentes con ellos y redacta al final una rúbrica alineada está ejecutando exactamente esta capacidad, con sus dos caras.

El uso instrumental de herramientas externas, tercera de las cuatro, convierte el lenguaje en acción efectiva. Chowa y colaboradores (2025) describen esta transición como el paso de modelos que solo generan texto a sistemas que ejecutan código, consultan bases de datos o navegan por la web para completar una tarea. En educación superior habilita agentes que consultan el sistema de gestión del aprendizaje, verifican una fecha de entrega real o recuperan bibliografía actualizada, en lugar de limitarse a lo que el modelo memorizó durante su entrenamiento.

Queda la memoria, que es la que más transforma la relación pedagógica y también la que peor se comprende. Zhang y colaboradores (2025) sistematizan el asunto en la primera revisión dedicada por completo a los mecanismos de memoria de estos sistemas, y distinguen la memoria interna del contexto, la que se recupera de un almacén externo y la que el propio agente decide escribir para su uso futuro. El origen técnico de la segunda está bien datado, en los modelos que consultan un corpus masivo en el momento de generar en lugar de confiar solo en sus parámetros (Borgeaud et al., 2022). La tercera es todavía terreno abierto, con propuestas de memoria organizada por el propio agente que apenas empiezan a evaluarse (Xu et al., 2025).

Un agente que recuerda las dificultades de un estudiante a lo largo de un semestre puede ajustar su intervención de un modo que ningún asistente sin memoria persistente iguala. Esa misma capacidad es la que convierte el expediente conversacional del estudiante en un archivo que alguien custodia, con las preguntas de privacidad y gobernanza de datos que abordamos en el capítulo 17. Conviene decidir esto antes de contratar, no después.



1.4 El bucle agéntico: percibir, planificar, actuar y evaluar

Estas cuatro capacidades no operan por separado. Se articulan en un ciclo que la literatura describe con bastante consistencia, con ligeras variaciones de nomenclatura, percibir, planificar, actuar y evaluar. La planificación es el eje que sostiene el resto, y de ella depende que el agente descomponga la tarea en pasos verificables en lugar de intentar resolverla de una vez (Huang et al., 2024). Masterman y colaboradores (2024) observan que los diseños que definen roles con claridad y refinan de forma iterativa sus salidas rinden mejor que los que resuelven en un único paso.

El ciclo comienza con la percepción. El agente recoge información relevante del entorno, la última entrega de un estudiante, un mensaje en el foro de la asignatura o una consulta directa. Con esa información entra en la planificación, donde descompone el objetivo en pasos concretos y ordenados. **La actuación ejecuta esos pasos, con frecuencia recurriendo a las herramientas externas del apartado anterior.** La evaluación cierra el bucle, comprueba si el resultado satisface el objetivo original y, cuando no lo hace, reinicia el proceso con la información adicional generada.

Ese cierre vale precisamente por lo que habilita, la corrección. Un modelo conversacional genera una respuesta y termina, sin verificar si resolvió el problema de quien preguntaba. Un agente que evalúa su propia salida puede detectar que la primera aproximación fue insuficiente y ensayar otra sin que el humano se lo indique. Chowa y colaboradores (2025) documentan mejoras apreciables en el anclaje al entorno y en la recuperación ante fallos cuando la orquestación entre razonamiento y acción sigue este patrón.

La honestidad epistémica exige una advertencia sobre los límites del bucle. Funciona razonablemente bien en tareas acotadas y se degrada en tareas largas y abiertas, donde el error de un paso contamina los siguientes, un efecto que Plaat y colaboradores (2026) identifican como el obstáculo persistente del razonamiento de varios pasos. A ello se añade un problema de comprobación. Xu (2026) señala que la reproducibilidad de estos sistemas sigue limitada por el no determinismo del modelo subyacente y por la variabilidad de las herramientas que invoca, y Arunkumar y colaboradores (2026) constatan que las prácticas de evaluación del campo todavía no permiten comparar arquitecturas de forma fiable.

Trasladado al aula, el diagnóstico se vuelve operativo. Un agente puede sostener con solvencia una tutoría de treinta minutos sobre un concepto acotado. Conviene vigilar de cerca su desempeño en tareas de acompañamiento que se extienden durante semanas, y conviene sobre todo no delegar en él la decisión final sobre el aprendizaje de nadie mientras esa fiabilidad no se pueda medir.

FIGURA
1.3

CAPÍTULO 1

El bucle agéntico

Percibir, planificar, actuar y evaluar no son cuatro fases aisladas, sino una disciplina de revisión continua



IDEA CLAVE El rasgo decisivo es el retorno: si el resultado no basta, el sistema revisa el plan.

Figura 1.3. El bucle agéntico: percibir, planificar, actuar y evaluar. Elaboración propia a partir de Huang et al. (2024), Masterman et al. (2024) y Chowa et al. (2025).

1.5 De los sistemas basados en reglas a las arquitecturas agénticas: qué transición estamos viviendo

La definición de agente que hoy se recicla en tantos folletos comerciales se formuló en 1995. La investigación en sistemas multiagente ya había delimitado entonces qué significaba percibir un entorno y actuar sobre él (Wooldridge y Jennings, 1995). Entre aquella formulación y los agentes contemporáneos median tres décadas de sistemas expertos con árboles de decisión codificados a mano. La idea es antigua. **Lo nuevo es el sustrato que permite ejecutarla a un coste razonable.**

El giro más reciente vuelve, curiosamente, al problema original de aquella tradición. Guo y colaboradores (2024) y Li y colaboradores (2024) documentan, en dos revisiones ampliamente citadas, cómo la investigación se ha desplazado del agente individual hacia la coordinación entre varios. Reaparecen entonces los asuntos que ya preocupaban a la ingeniería de agentes de los años noventa, división del trabajo, comunicación y resolución de conflictos. El capítulo 5 desarrolla esa vuelta con detalle.

Esta transición no ha sido instantánea ni completa, y conviene resistir la tentación de narrarla como una ruptura limpia. Durante años convivieron sistemas basados en reglas con primeras aproximaciones de agentes reactivos, y buena parte de la infraestructura educativa universitaria sigue operando hoy con automatización tradicional, sin ningún componente agéntico. Lo que ha cambiado en los últimos dos años es la disponibilidad de una alternativa que antes no existía a un precio asumible.

¿Qué hace posible el salto? El rendimiento de un sistema agéntico depende de cómo se combinan planificación, memoria, reflexión y uso de herramientas para perseguir objetivos de varios pasos, más que de la potencia bruta del modelo de lenguaje que lo sostiene (Huang et al., 2024, Zhang et al., 2025, Xu, 2026). La precisión importa a quien diseña una implementación institucional, asunto del capítulo 18. La orquestación y el diseño del sistema, y no solo el modelo elegido, determinan si un agente educativo alcanza la fiabilidad que una universidad puede permitirse.



1.6 Clarificación terminológica: agente, agenticidad, IA agéntica

Cerramos con una precisión terminológica que evitará ambigüedades en el resto del volumen. Empleamos tres términos que se solapan en el uso corriente y que conviene distinguir con cierto rigor.

Llamamos *agente*, siguiendo la formulación clásica de Wooldridge y Jennings (1995), a un sistema que percibe su entorno, procesa esa información mediante algún mecanismo de razonamiento y actúa sobre dicho entorno para alcanzar un objetivo. Esta definición, acuñada casi tres décadas antes de que existieran los modelos de lenguaje grande, sigue siendo sorprendentemente útil, aunque en el capítulo 2 mostraremos cómo su aplicación a sistemas probabilísticos introduce complicaciones que sus autores no anticiparon.

Entendemos por *agenticidad* la propiedad gradual, no binaria, que un sistema exhibe en mayor o menor medida según combine autonomía, proactividad, capacidad de planificación y memoria persistente. Un sistema se sitúa en un punto de un espectro. Las taxonomías técnicas disponibles ordenan ese espectro por arquitectura antes que por umbral, lo que respalda esta lectura gradual (Xu, 2026, Arunkumar et al., 2026). La gradualidad evita el error, frecuente en la divulgación tecnológica, de etiquetar como agente cualquier producto que incorpore un asistente conversacional con alguna función añadida.

Reservamos la expresión *IA agéntica* para una arquitectura contemporánea: un modelo de lenguaje grande integrado en un bucle de software que planifica, invoca herramientas y conserva memoria entre interacciones. La capacidad agéntica surge de esa construcción, no del modelo aislado. El matiz sostiene la distinción transversal del manual. Podemos delegar autonomía operativa dentro de ciertos límites. La agencia epistémica exige otra cautela, porque compromete la formación del juicio propio del estudiante. El capítulo siguiente examina esa diferencia con más profundidad.

NOTA**TRES TÉRMINOS, TRES NIVELES**

Agente: sistema que percibe, razona y actúa sobre su entorno para alcanzar un objetivo (Wooldridge y Jennings, 1995). Agenticidad: propiedad gradual que combina autonomía, proactividad, planificación y memoria. IA agéntica: arquitectura que integra un modelo de lenguaje grande en un bucle de software capaz de planificar, usar herramientas y recordar.

Síntesis del capítulo

Cuatro capacidades separan al agente de la asistencia conversacional: autonomía operativa, planificación de varios pasos, uso de herramientas externas y memoria persistente. Juntas forman un bucle de percepción, planificación, acción y evaluación que permite corregir el rumbo sin pedir aprobación humana en cada paso.

Ese bucle pierde fiabilidad en tareas largas y abiertas. Tampoco disponemos todavía de prácticas de evaluación que permitan comparar arquitecturas con suficiente solvencia. La asimetría probatoria es clara: la evidencia técnica crece deprisa, mientras que la evidencia pedagógica en educación superior sigue siendo escasa. Los datos disponibles favorecen los flujos híbridos con supervisión docente.

La terminología queda fijada. Un agente percibe, razona y actúa. La agenticidad es una propiedad gradual. La IA agéntica describe una arquitectura, no una cualidad del modelo aislado. Con esta base, el capítulo siguiente examina en qué sentido cabe llamar agente a un sistema estadístico.

[Volver al índice](#)

2

El problema ontológico del agente

De este capítulo se incluye únicamente el apartado 2.6, el que enuncia la distinción sobre la que se apoya el resto de la obra. Los seis apartados restantes recorren la genealogía del concepto, de la cibernética a los sistemas multiagente clásicos, y quedan en el libro.

UN APARTADO DE SIETE

2.6 Autonomía operativa y agencia epistémica: qué puede delegarse y qué no

Llegamos al eje que sostiene, de manera transversal, todo este manual. Proponemos distinguir entre dos tipos de agencia que la conversación pública suele confundir bajo una misma palabra. Llamamos *autonomía operativa* a la capacidad de un sistema para ejecutar tareas, secuenciar pasos y coordinar herramientas sin supervisión constante. Llamamos *agencia epistémica* a la capacidad de formar juicio propio, evaluar críticamente una evidencia y construir comprensión sobre un problema. La primera describe una propiedad del sistema. La segunda describe una capacidad de la persona, y por eso la delegación afecta a cada una de manera asimétrica.

La autonomía operativa puede delegarse sin mayor riesgo conceptual. Que un sistema programe las fechas de un plan de estudio, agrupe estudiantes según su progreso o genere un primer borrador de rúbrica no compromete nada esencial del proceso formativo, siempre que un docente revise el resultado. La agencia epistémica no puede delegarse sin coste, porque delegarla equivale a delegar la formación misma del juicio que la universidad existe para cultivar.

TABLA 2.1 · LAS DOS AGENCIAS: QUÉ SE DELEGA Y QUÉ NO

	Autonomía operativa	Agencia epistémica
Qué nombra	Capacidad de ejecutar tareas , secuenciar pasos y coordinar herramientas sin supervisión constante.	Capacidad de formar juicio propio , evaluar críticamente una evidencia y construir comprensión.
De quién es	Propiedad <i>del sistema</i> .	Capacidad <i>de la persona</i> .
Ejemplos	Programar las fechas de un plan de estudio, agrupar estudiantes según su progreso, generar un primer borrador de rúbrica.	Decidir qué merece ser creído, sostener una interpretación, responder por una conclusión.
¿Puede delegarse?	Sí , siempre que un docente revise el resultado.	No sin coste : delegarla equivale a delegar la formación del juicio que la universidad existe para cultivar.

Marco original de este manual. La escala de niveles de autonomía procede de Feng et al. (2025).

CINCO NIVELES DE AUTONOMÍA, SEGÚN EL PAPEL QUE CONSERVA LA PERSONA

Operador	Colaborador	Consultor	Aprobador	Observador
----------	-------------	-----------	-----------	------------

La persona **hace**La persona **mira**

Feng et al. (2025). El capítulo 7 del manual desarrolla los cinco niveles y su matriz de decisión.

Feng y colaboradores (2025) aportan aquí un instrumento que conviene incorporar. Sostienen que el grado de autonomía de un agente es una decisión de diseño separable de su capacidad técnica, y proponen cinco niveles definidos por el papel que conserva la persona, operador, colaborador, consultor, aprobador y observador. La escala sustituye una pregunta binaria por una graduada, en qué papel queremos que queden el docente y el estudiante al terminar la tarea.

La distinción no es una intuición sin respaldo, aunque la evidencia resulta más matizada de lo que sugieren los titulares. Gerlich (2025), con seiscientos sesenta y seis participantes, encuentra una correlación negativa significativa entre el uso frecuente de herramientas de inteligencia artificial y las capacidades de pensamiento crítico, mediada por lo que la literatura denomina *descarga cognitiva*. El trabajo se cita con profusión y merece leerse con cuidado, porque es correlacional, se apoya en medidas autoinformadas y su muestra no es universitaria, de modo que autoriza a sospechar de un mecanismo sin permitir afirmar una causa. Li y colaboradores (2026) permiten afinar al sintetizar sesenta y siete estudios empíricos, y condicionan el efecto al encuadre pedagógico, con ganancias bajo *andamiaje* y descensos en uso no estructurado.

Wang y colaboradores (2026), en una revisión de alcance sobre ciento veintitrés estudios, describen un patrón doble. Hay configuraciones que sostienen la agencia del estudiante, asociadas a la autorregulación y a la competencia para interpretar la retroalimentación, y configuraciones que la erosionan, asociadas a la descarga cognitiva y a la aceptación acrítica. La comprobación más directa que conocemos la aporta el cuasiexperimento de Hong y colaboradores (2025). Doscientos cuarenta estudiantes trabajaron doce semanas en ciclos de escritura donde el sistema asumía las tareas de orden inferior y el estudiante conservaba el análisis, la evaluación y la reflexión. Ese grupo mejoró de forma significativa en pruebas normalizadas de pensamiento crítico. Delegar no perjudicó. Delegar lo que no había que delegar es lo que perjudica, aunque convenga esperar réplicas en otras materias antes de generalizar un solo diseño en un único dominio.

Vendrell y Johnston (2026) trasladan esta lógica a principios de diseño, y hablan expresamente de agencia epistémica como proceso que el diseño debe proteger. Entre sus principios figura el de preservar la fricción cognitiva, mantener de manera deliberada el esfuerzo que el sistema podría ahorrar cuando ese esfuerzo es justamente lo que produce el aprendizaje. Heersmink y colaboradores (2024) añaden el matiz que cierra el argumento, porque los rasgos que hacen atractivo a un modelo, su fluidez y la sensación de flujo interaccional, dificultan calibrar la confianza en lo que produce. Confiar demasiado poco desperdicia una herramienta útil. Confiar demasiado entrega a un sistema estadístico una función que solo un juicio humano formado puede ejercer con responsabilidad.

Esta distinción ofrece un criterio de decisión, no una frontera fija entre tareas. Ante cualquier delegación proponemos preguntar qué perdería el estudiante si dejara de practicar aquello que el sistema puede hacer por él. El capítulo 7 convierte este criterio en el marco de autonomía pedagógicamente restringida.

FIGURA
2.4

CAPÍTULO 2

Qué puede delegarse y qué no

La autonomía operativa admite delegación. La agencia epistémica define la responsabilidad intelectual de la persona

DELEGABLE CON LÍMITES**Autonomía operativa**

Acciones que amplían la capacidad de trabajo.

- ✓ Secuenciar tareas
- ✓ Organizar calendarios
- ✓ Dar formato
- ✓ Buscar información

NO DELEGABLE**Agencia epistémica**

Capacidades que sostienen la autoría y la responsabilidad.

- ✗ Ejercer juicio crítico
- ✗ Evaluar evidencia
- ✗ Construir comprensión
- ✗ Asumir criterio ético

IDEA CLAVE Delegar una operación no equivale a ceder el juicio sobre qué creer, por qué creerlo o qué consecuencias aceptar.

Figura 2.4. Qué puede delegarse y qué no: autonomía operativa frente a agencia epistémica. Marco original de este manual, con los niveles de autonomía de Feng et al. (2025) como escala de referencia.

Los dos ensayos que han medido el aprendizaje de manera más directa no evalúan agentes, sino inteligencia artificial generativa conversacional. Los traemos aquí con esa salvedad explícita, porque un sistema con más autonomía amplificaría previsiblemente lo que un asistente reactivo ya produce, aunque esto último sea una conjetura y no un hallazgo.

Bastani y colaboradores (2025) realizaron un ensayo aleatorizado con cerca de mil estudiantes de secundaria en Turquía, repartidos en tres condiciones: acceso a una interfaz equivalente a GPT-4 sin restricciones, acceso a una versión tutorizada cuyas instrucciones fueron diseñadas con el profesorado para guiar con pistas en lugar de entregar la respuesta, y un grupo de control sin herramienta. Durante la práctica, los dos grupos con acceso rindieron muy por encima del control, un 48 % y un 127 % mejor respectivamente. En el examen posterior, ya sin acceso a nada, el grupo que había usado la interfaz sin salvaguardas rindió un 17 % peor que el control. Las salvaguardas eliminaron ese efecto negativo sin llegar a producir uno positivo. Los autores interpretan que los estudiantes emplearon el modelo como muleta y no comprobaron sus errores. Es secundaria y es matemáticas, no educación superior, y conviene retenerlo. Aun así, es el ensayo aleatorizado de mayor tamaño publicado hasta la fecha sobre esta cuestión, y su hallazgo apunta exactamente adonde apunta este manual.

Fan y colaboradores (2025) trabajaron ya en educación superior, con ciento diecisiete estudiantes universitarios en una tarea de escritura y cuatro condiciones: apoyo de un modelo conversacional, apoyo de un experto humano, herramientas de analítica de la escritura y ningún apoyo adicional. El grupo con apoyo del modelo obtuvo la mayor mejora en la puntuación del ensayo, y sin embargo su ganancia de conocimiento y su transferencia no difirieron de manera significativa de las de los demás grupos. Los procesos de autorregulación sí difirieron, con motivación equivalente. Los autores denominan pereza metacognitiva a ese patrón. La muestra es pequeña, la tarea única y el contexto uno solo.

Puestos uno junto a otro, ambos apuntan en la misma dirección y con una precisión que conviene no perder: el producto mejora y el aprendizaje no. En el primer caso, con acceso irrestricto, el aprendizaje empeora. **Y en los dos, la variable decisiva resultó ser el diseño de las salvaguardas, no la potencia del modelo.** Es la formulación empírica más nítida de la asimetría que sostiene este apartado: la ejecución se delegó y mejoró el resultado, el juicio se delegó y desapareció el aprendizaje.

TABLA 2.2 · QUÉ DICE LA EVIDENCIA, Y QUÉ NO PERMITE AFIRMAR

Estudio	Alcance	Qué encuentra	Qué no autoriza a afirmar
Bastani et al. (2025)	Ensayo aleatorizado, cerca de 1.000 estudiantes	Con acceso sin salvaguardas, 17 % peor en el examen posterior . Las salvaguardas eliminan el efecto negativo.	Es secundaria y matemáticas, no educación superior.
Fan et al. (2025)	117 estudiantes universitarios	Mejor puntuación del ensayo, sin diferencias en ganancia de conocimiento ni transferencia.	Muestra pequeña, una sola tarea, un solo contexto.
Gerlich (2025)	666 participantes	Correlación negativa entre uso frecuente y pensamiento crítico, mediada por <i>descarga cognitiva</i> .	Correlacional y con muestra no universitaria: sugiere un mecanismo, no una causa.
Hong et al. (2025)	Cuasiexperimento, 240 estudiantes, 12 semanas	El grupo que delegó la ejecución y conservó el juicio mejoró de forma significativa .	Un solo diseño en un único dominio.

Elaboración propia a partir de los estudios citados en el apartado 2.6.

APARATO DE LA OBRA

Referencias

28 entradas, verificadas contra los registros de Crossref y DataCite.

- Arunkumar, V., G. R., G., y Buyya, R. (2026). Agentic artificial intelligence (AI): Architectures, taxonomies, and evaluation of large language model agents. *arXiv*. <https://doi.org/10.48550/arXiv.2601.12560>
- Borgeaud, S., Mensch, A., Hoffmann, J., Cai, T., Rutherford, E., Millican, K., van den Driessche, G., Lespiau, J.-B., Damoc, B., Clark, A., de las Casas, D., Guy, A., Menick, J., Ring, R., Hennigan, T., Huang, S., Maggiore, L., Jones, C., Cassirer, A., y Sifre, L. (2022). Improving language models by retrieving from trillions of tokens. *Proceedings of the 39th International Conference on Machine Learning*, 162, 2206-2240. <https://doi.org/10.48550/arXiv.2112.04426>
- Chowa, S. S., Alvi, R., Rahman, S. S., Rahman, M. A., Raiaan, M. A. K., Islam, M. R., Hussain, M., y Azam, S. (2025). From language to action: A review of large language models as autonomous agents and tool users. *Artificial Intelligence Review*, 59. <https://doi.org/10.1007/s10462-025-11471-9>
- Córdova-Esparza, D.-M. (2025). AI-powered educational agents: Opportunities, innovations, and ethical challenges. *Information*, 16(6), 469. <https://doi.org/10.3390/info16060469>
- Eager, B., y Brunton, R. (2023). Prompting higher education towards AI-augmented teaching and learning practice. *Journal of University Teaching and Learning Practice*, 20(5). <https://doi.org/10.53761/1.20.5.02>
- Essel, H. B., Vlachopoulos, D., Essuman, A. B., y Amankwa, J. O. (2023). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence*, 6, 100198. <https://doi.org/10.1016/j.caeai.2023.100198>
- Graesser, A. C., Chipman, P., Haynes, B. C., y Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612-618. <https://doi.org/10.1109/TE.2005.856149>
- Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., y Zhang, X. (2024). Large language model based multi-agents: A survey of progress and challenges. *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 8048-8057. <https://doi.org/10.48550/arXiv.2402.01680>
- Huang, X., Liu, W., Chen, X., Wang, X., Wang, H., Lian, D., Wang, Y., Tang, R., y Chen, E. (2024). Understanding the planning of LLM agents: A survey. *arXiv*. <https://doi.org/10.48550/arXiv.2402.02716>
- Kostopoulos, G., Gkamas, V., Rigou, M., y Kotsiantis, S. (2025). Agentic AI in education: State of the art and future directions. *IEEE Access*, 13, 177467-177491. <https://doi.org/10.1109/ACCESS.2025.3620473>
- Laato, S., Morschheuser, B., Hamari, J., y Björne, J. (2023). AI-assisted learning with ChatGPT and large language models: Implications for higher education. *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)*, 226-230. <https://doi.org/10.1109/ICALT58122.2023.00072>
- Li, X., Wang, S., Zeng, S., Wu, Y., y Yang, Y. (2024). A survey on LLM-based multi-agent systems: Workflow, infrastructure, and challenges. *Vicinearth*, 1. <https://doi.org/10.1007/s44336-024-00009-2>
- Masterman, T., Besen, S., Sawtell, M., y Chao, A. (2024). The landscape of emerging AI agent architectures for reasoning, planning, and tool calling: A survey. *arXiv*. <https://doi.org/10.48550/arXiv.2404.11584>
- Plaat, A., Wong, A., Verberne, S., Broekens, J., van Stein, N., y Bäck, T. (2026). Multi-step reasoning with large language models, a survey. *ACM Computing Surveys*, 58, 1-35. <https://doi.org/10.1145/3774896>
- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., y Wen, J. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 186345. <https://doi.org/10.1007/s11704-024-40231-1>
- Wooldridge, M., y Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2), 115-152. <https://doi.org/10.1017/S0269888900008122>
- Xu, B. (2026). AI agent systems: Architectures, applications, and evaluation. *arXiv*. <https://doi.org/10.48550/arXiv.2601.01743>

- Xu, W., Liang, Z., Mei, K., Gao, H., Tan, J., y Zhang, Y. (2025). A-MEM: Agentic memory for LLM agents. arXiv. <https://doi.org/10.48550/arXiv.2502.12110>
- Yigci, D., Eryilmaz, M., Yetisen, A. K., Tasoglu, S., y Ozcan, A. (2025). Large language model-based chatbots in higher education. *Advanced Intelligent Systems*, 7, 2400429. <https://doi.org/10.1002/aisy.202400429>
- Zhang, Z., Dai, Q., Bo, X., Ma, C., Li, R., Chen, X., Zhu, J., y Dong, Z. (2025). A survey on the memory mechanism of large language model-based agents. *ACM Transactions on Information Systems*, 43, 1-47. <https://doi.org/10.1145/3748302>
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakci, Ö., y Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., y Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489-530. <https://doi.org/10.1111/bjet.13544>
- Feng, K. J. K., McDonald, D. W., y Zhang, A. X. (2025). Levels of autonomy for AI agents (Ensayo 25-15). Knight First Amendment Institute at Columbia University. <https://doi.org/10.48550/arXiv.2506.12469>
- Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- Heersmink, R., de Rooij, B., Clavel Vázquez, M. J., y Colombo, M. (2024). A phenomenology and epistemology of large language models: Transparency, trust, and trustworthiness. *Ethics and Information Technology*, 26(3), Artículo 41. <https://doi.org/10.1007/s10676-024-09777-3>
- Hong, H., Vate-U-Lan, P., y Viriyavejakul, C. (2025). Cognitive offload instruction with generative AI: A quasi-experimental study on critical thinking gains in English writing. *Forum for Linguistic Studies*, 7(7). <https://doi.org/10.30564/fls.v7i7.10072>
- Li, C., Cui, H., y Hagedorn, L. S. (2026). The cognitive impact of ChatGPT in higher education: A systematic review of critical and creative thinking outcomes. *Computers and Education: Artificial Intelligence*, 10, 100571. <https://doi.org/10.1016/j.caeai.2026.100571>
- Wang, G., Wang, W., Yang, D., y Ren, J. (2026). Generative AI, cognitive offloading, and learner agency in higher education: A scoping review. *Behavioral Sciences*, 16(7), 1150. <https://doi.org/10.3390/bs16071150>

HASTA AQUÍ LA MUESTRA

IA agéntica en educación superior

El manual continúa con las arquitecturas y los patrones de coordinación, el diseño pedagógico con agentes, las aplicaciones documentadas en educación superior y la parte de confiabilidad, evaluación y gobernanza. Cierra con seis anexos de trabajo y una bibliografía comentada, verificada referencia por referencia.

20

CAPÍTULOS

7

ANEXOS

73

FIGURAS

3

CASOS
DOCUMENTADOS

SE PUBLICA EL JUEVES 3 DE SEPTIEMBRE DE 2026